Jetstream - In production & supporting science

Available to **everyone** (including you, engineers! And you too, anyone with a biological collection)

Craig A. Stewart, Executive Director, IU Pervasive Technology Institute, Jetstream P.I.
David Y. Hancock, Program Director, Advanced Cyberinfrastructure, IU Pervasive Technology Institute
Jennifer Laherty, Associate Librarian, Head of Sciences, IU Libraries

Outline

- Overview of Jetstream including current status
- Info on orchestrated use of Jetstream and new cool tools (e.g. Docker, Mesos, Kubernetes, etc.)
- Info of interest to researchers working in engineering
- Example of Jetstream aiding a research area not typically supported by XSEDE: Imago, a tool for managing biological collections

Flickr Canned Muffins 0365_092 – CC 2.0 Licensed

Jetstream
http://jetstream-cloud.org/

funded by the National Science Foundation
Award #ACI-1445604

Jetstream - Expanding NSF XD's reach and impact

Around 350,000 researchers, educators, & learners received NSF support in 2015

- Less than 2% completed a computation, data analysis, or visualization task on XD/XSEDE program resources
- Less than 4% had an XSEDE Portal account
- 70% of researchers surveyed* claimed to be resource constrained

Why aren't they using XD/XSEDE systems?

- Activation energy is pretty high
- HPC resources are scarce and not well-matched to their needs
- They just don't need *that much* capability

* https://www.xsede.org/xsede-nsf-release-cloud-survey-report

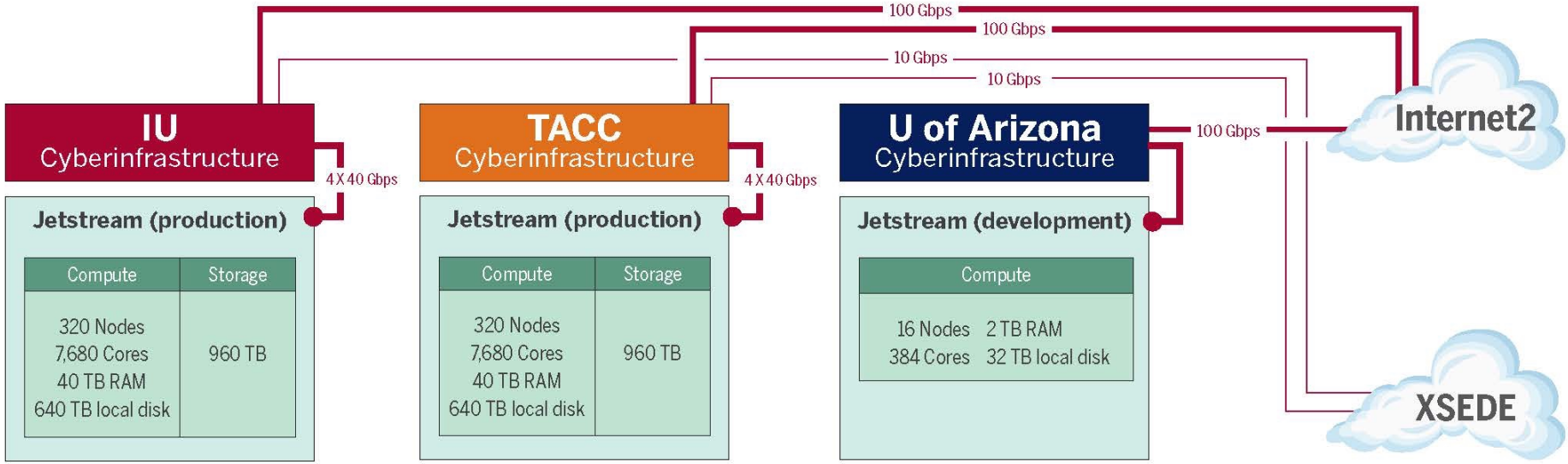
What is Jetstream?

- NSF's first production cloud facility
- Part of the NSF eXtreme Digital (XD) program
- Provides on-demand *interactive* computing and analysis
- Enables *configurable* environments and architectures
- User-friendly, widely accessible cloud environment
- User-selectable library of preconfigured virtual machines

Jetstream
http://jetstream-cloud.org/

funded by the National Science Foundation
Award #ACI-1445604

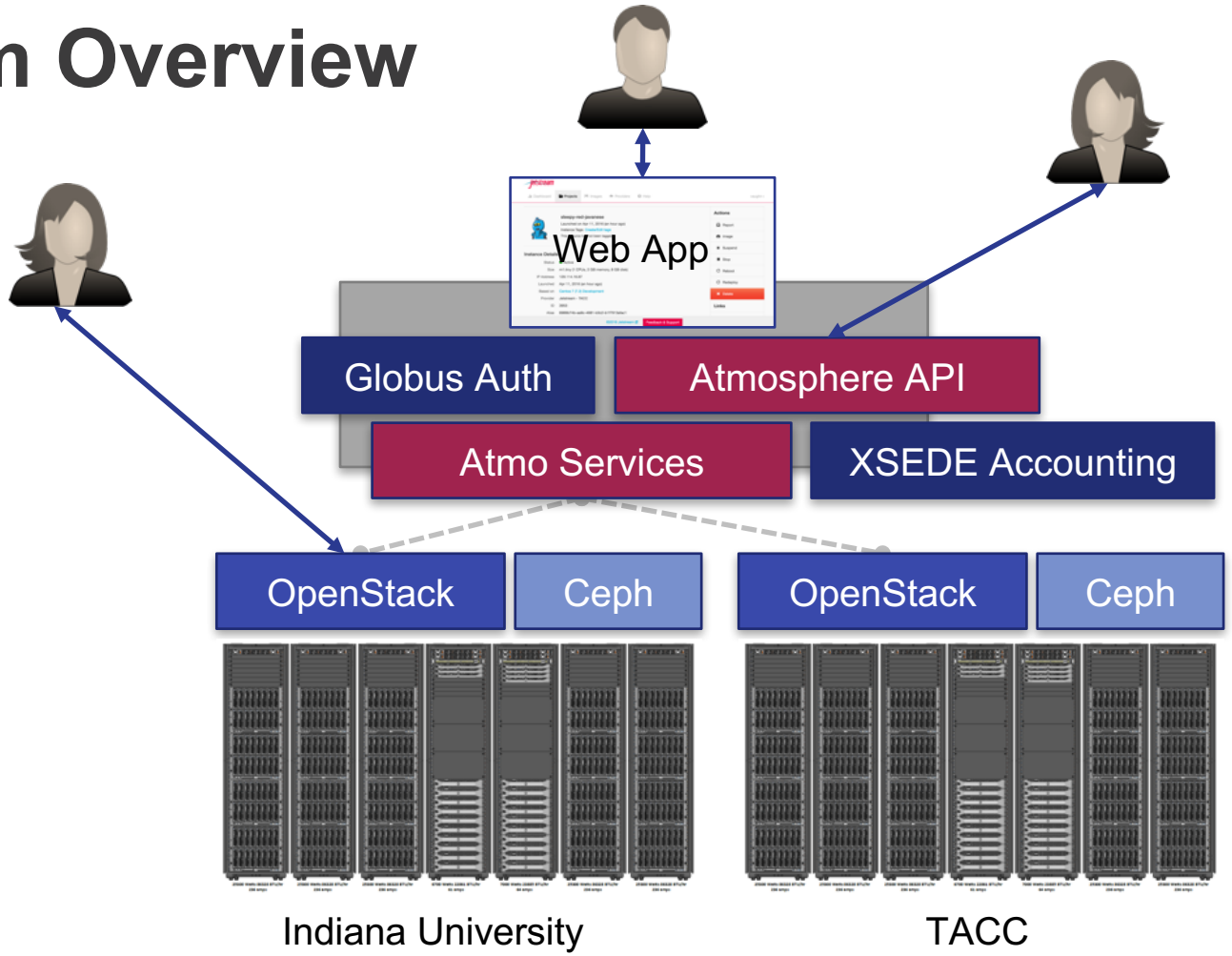
Jetstream System Overview



Jetstream
http://jetstream-cloud.org/

NSF funded by the National Science Foundation
Award #ACI-1445604

Platform Overview



Hardware and Instance ”Flavors”

VM Host Configuration

- Dual Intel E-2680v3 “Haswell”
- 24 physical cores/node @ 2.5 GHz (Hyperthreading on)
- 128 GB RAM
- Dual 1 TB local disks
- 10GB dual uplink NIC
- Using KVM Hypervisor

Flavor	vCPUs	RAM	Storage	SU Per Hour
m.tiny	1	2	8	1
m.small	2	4	20	2
m.medium	6	16	60	6
m.large	10	30	120	10
m.xlarge	24	60	240	24
m.xxlarge	44	120	480	44

- Short-term storage comes as part of a launched instance
- Long-term storage mounted though block volumes
- Each user gets 10 volumes up to 500GB total storage
- Piloting object storage

Who uses Jetstream?

Researchers needing a handful of cores (1 to 44/vCPU)

Software creators and researchers needing to create their own customized virtual machines, containers, and workflows

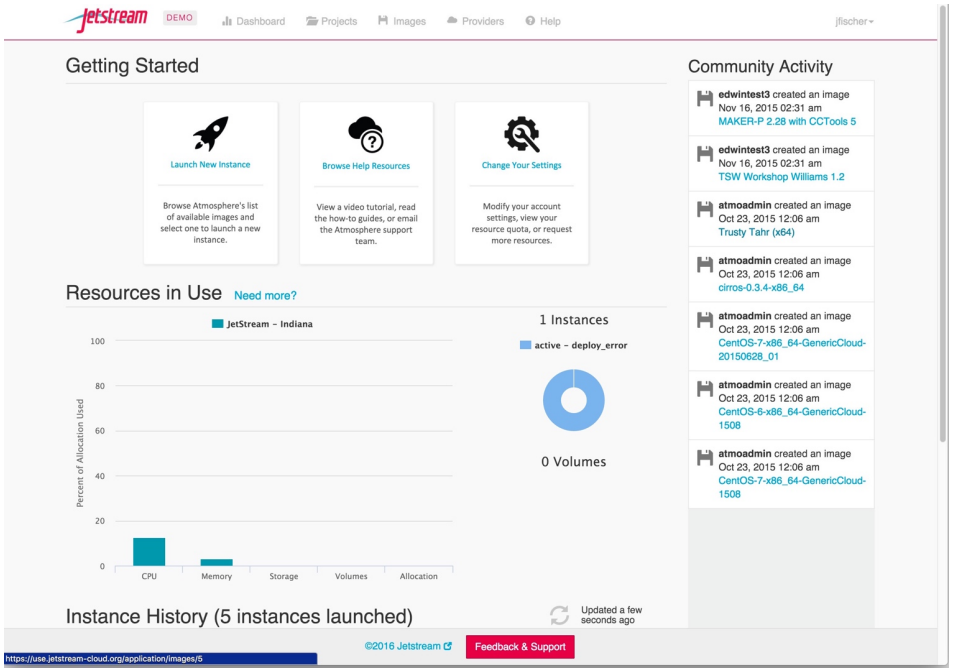
Science gateway creators using Jetstream as either the front end or processor for scientific workflows

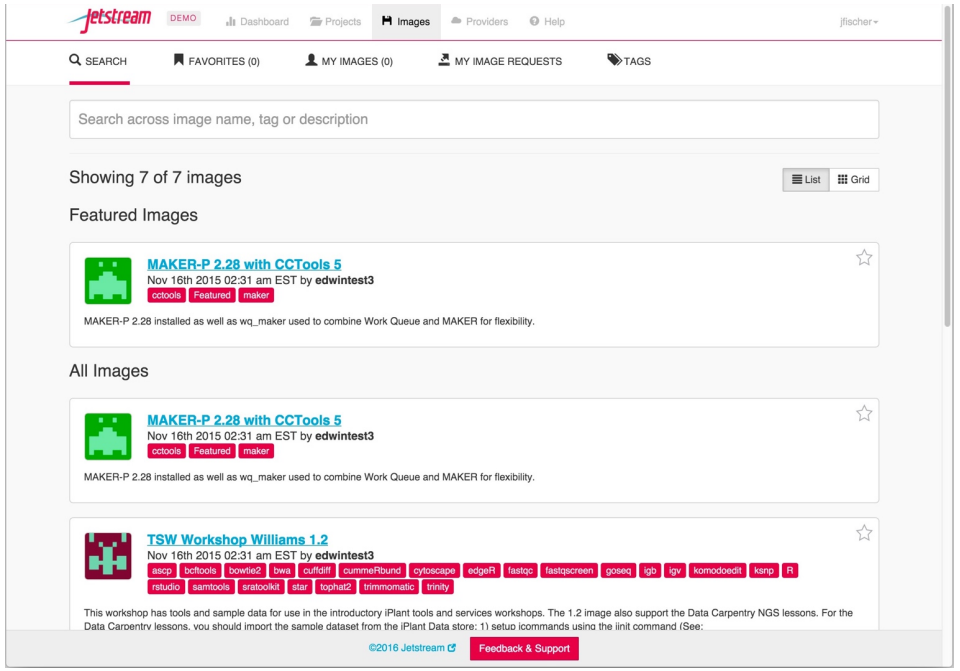
Educators teaching on a variety of subjects

Jetstream
http://jetstream-cloud.org/

NSF funded by the National Science Foundation
Award #ACI-1445604

The Jetstream Interface





Jetstream
http://jetstream-cloud.org/

NSF
funded by the National Science Foundation
Award #ACI-1445604

Methods of access

Primary methods of access

- Interactive user access via web interface with VNC/SSH
- Direct VNC/SSH to individual instances
- Direct OpenStack or Atmosphere API access
- API access enables Science Gateways and other *always on* services or *on demand* use cases; e.g. elastic compute techniques

Jetstream
http://jetstream-cloud.org/

NSF funded by the National Science Foundation
Award #ACI-1445604

You want to use Mesos, Docker, Kubernetes, whatever the latest and greatest new thing is? You want to run a science gateway?

- Two things keep happening to us:
  - We get asked “Do you support <new cool thing>?”
  - And then someone builds a VM image that makes use of <new cool thing>
- Several researchers are using Docker and Mesos already, and we are experimenting with Kubernetes.
- Orchestration engines make it easier to provision Jetstream resources behind gateways
- Supporting science gateways and encouraging orchestration engine use facilitates scaling applications and greated CPU usage

Jetstream
http://jetstream-cloud.org/

NSF funded by the National Science Foundation
Award #ACI-1445604

Science Domains Examples

Biology: iPlant and Galaxy VMs

Earth Science: NSIDC data analysis, EarthCube ECITE/CHORDS

Field Station Research: data collection and analysis tools to support data sharing and collaboration

Network Science: Network Workbench gateway and VMs

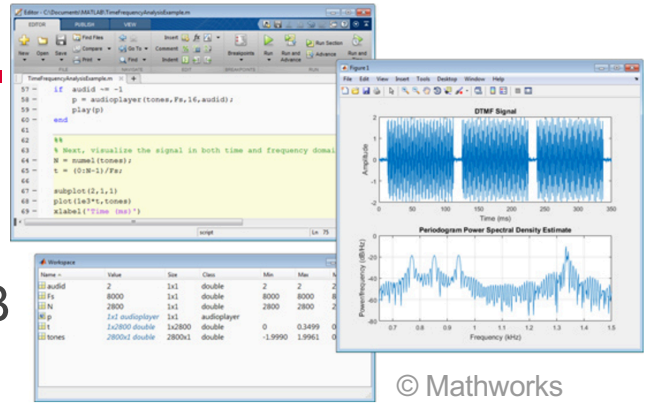
Social Sciences: VMs utilizing data from the Odum Institute (and others)

Computer Science/Cyberinfrastructure: RADICAL Tools, several education allocations

Whatever you do, probably… unless you run large scale MPI codes or need accelerators

funded by the National Science Foundation
Award #ACI-1445604

Jetstream for engineering researchers



© Mathworks

Matlab and the 52 standard toolkits are installed on Jetstream

You do **NOT** need to have a local license to use MATLAB on Jetstream

If you are a researcher, and MATLAB and one of the standard first 52 toolkits that come with MATLAB help you… **have at it!**

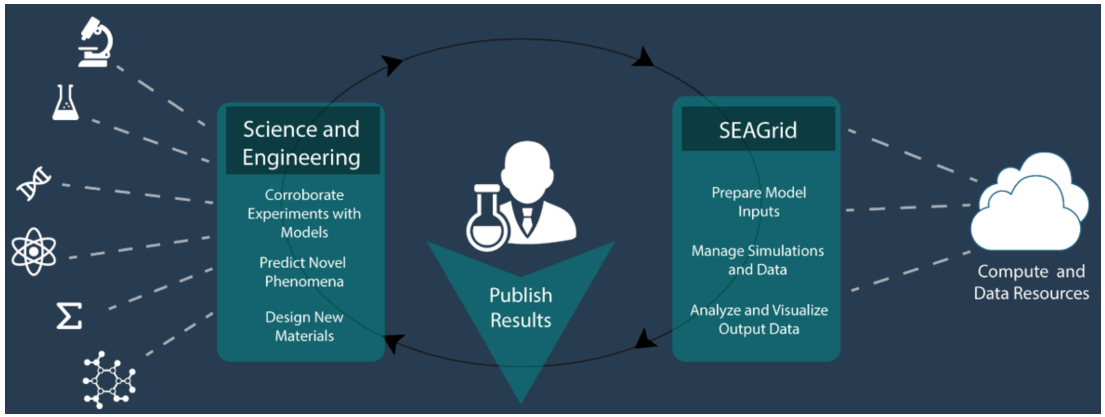
If you are an engineering researcher, and you need other tools... Let us know – we are happy to consider other requests

Jetstream
http://jetstream-cloud.org/

NSF funded by the National Science Foundation
Award #ACI-1445604

Science/Engineering Applications

- Using OpenStack Heat to create a PBS cluster
- Ansible is used to dynamically install/configure applications
- Leveraging Apache Airavata GFac to submit workloads
- Supports apps such as:
- Quantum Espresso
- Gromacs
- LAMMPS
- NWChem

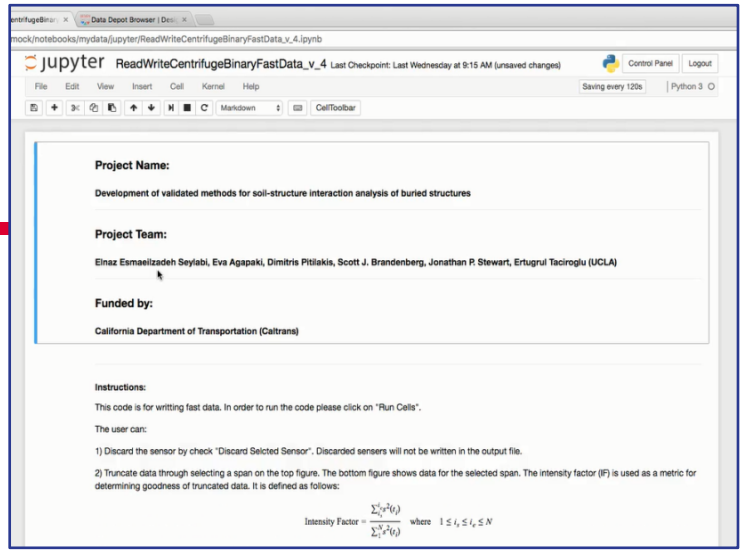


Jetstream
http://jetstream-cloud.org/

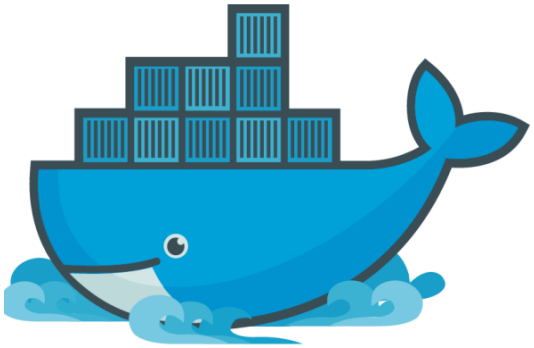
funded by the National Science Foundation
Award #ACI-1445604

DESIGNSAFE-CI

A NATURAL HAZARDS ENGINEERING COMMUNITY



- Shared-use research infrastructure funded by NSF to enable transformative research in natural hazards engineering
- Jobs run (via Agave) on HPC and Cloud systems
- VM sized applications packaged as Docker containers where appropriate for portability
- Using Jupyter with bursting capabilities to Jetstream to support high usage training



Jetstream
http://jetstream-cloud.org/

funded by the National Science Foundation
Award #ACI-1445604

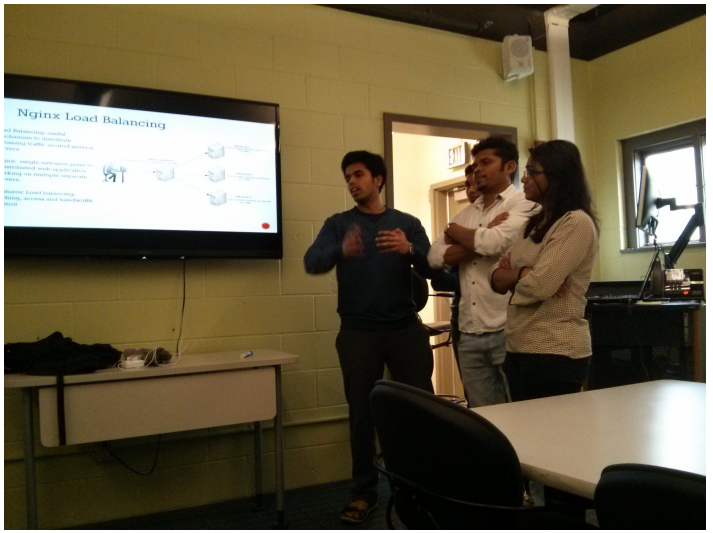
Supporting Education

Jetstream has been used in multiple IU Informatics Graduate Courses

- INFO 535 – Management, Access, and Use of Big and Complex Data
- INFO 590 – Topics in Informatics

BlueWaters Workflow Workshop

Multiple Research Data Alliance Workshops



Jetstream
http://jetstream-cloud.org/

NSF funded by the National Science Foundation
Award #ACI-1445604

An example of Jetstream enabling services for groups of scientists and citizens who never before would have thought of using an XSEDE-supported resource

Introducing Imago

For Physical Specimen Biocollections

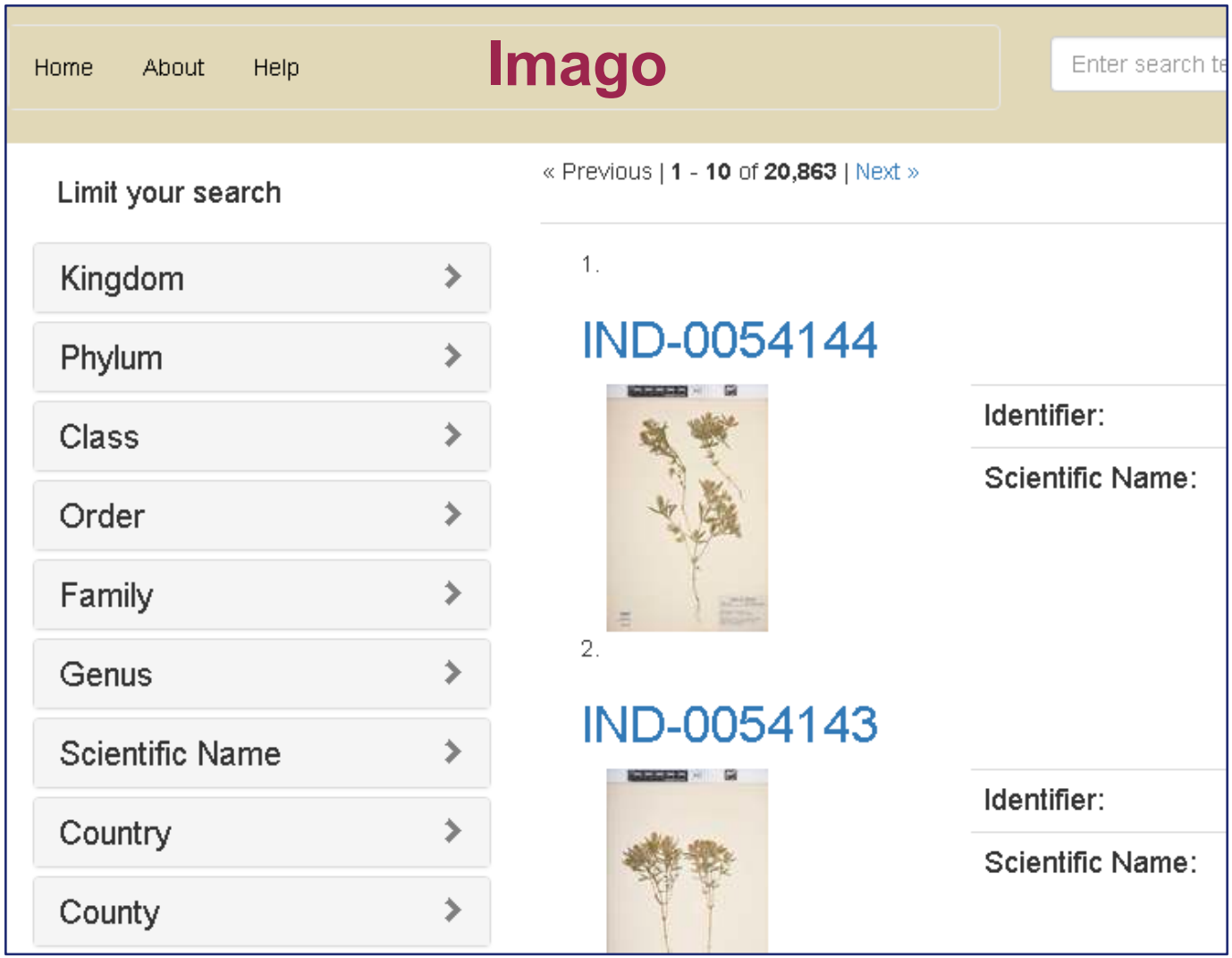
-Objects:
Digital Library Preservation Repository
2-D & 3-D images

-Metadata

2017
-Available on a VM in Jetstream

-Import/Export functions to facilitate workflows with NSF supported Specify and Symbiota

http://imago.indiana.edu/catalog



IND-0010840 Open Access

Descriptions

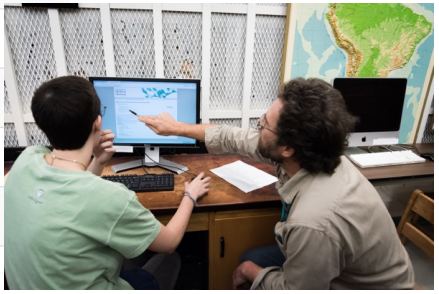
Attribute Name	Values
Identifier	http://purl.dlib.indiana.edu/iudl/herbarium/IND-0010840
Rights	Attribution-NonCommercial 3.0 United States
Collection code	herbarium
Catalog number	IND-0010840
Country	United States
State province	Indiana
County	Shelby
Kingdom	Plantae
Phylum	Magnoliophyta
Order	Liliales
Family	Melanthiaceae
Genus	Trillium
Specific epithet	recurvatum
Scientific name	Trillium recurvatum
Scientific name authorship	Beck
Basis of record	PreservedSpecimen

Last modified:

Download the full-sized image

0

Citations:
EndNote | Zotero | Mendeley



Journey: Metadata + Digital Images

Imago

Symbiota Promoting Bio-Collaboration

Specify Software Project

GBIF Global Biodiversity Information Facility

iDigBio Integrated Digitized Biocollections

EarthCube

USGS science for a changing world

Biodiversity Information Serving Our Nation (BISON) - Explore & download U.S. species occurrence data & maps

Impact on Research

ELSEVIER

Palaeogeography, Palaeoclimatology, Palaeoecology

Volume 449, 1 May 2016, Pages 227–235

Exploration of marine mammal paleogeography in the Northern Hemisphere over the Cenozoic using beta diversity

Carlos Mauricio Peredo[a], Mark D. Uhen[b]

Show more

http://dx.doi.org/10.1016/j.palaeo.2016.02.034

Highlights

- Beta diversity is used to assess use of distributional passageways by marine mammals.
- The Strait of Gibraltar is used by Cetacea and Sirenia to exit the Tethys Sea in the Eocene.
- The Strait of Gibraltar is used to reinvade the Mediterranean Sea in the Pliocene.
- The Central American Seaway plays only a minor role in marine mammal distribution.
- Modern North Atlantic–North Pacific Ocean similarity is established in the Quaternary via the Arctic Ocean.

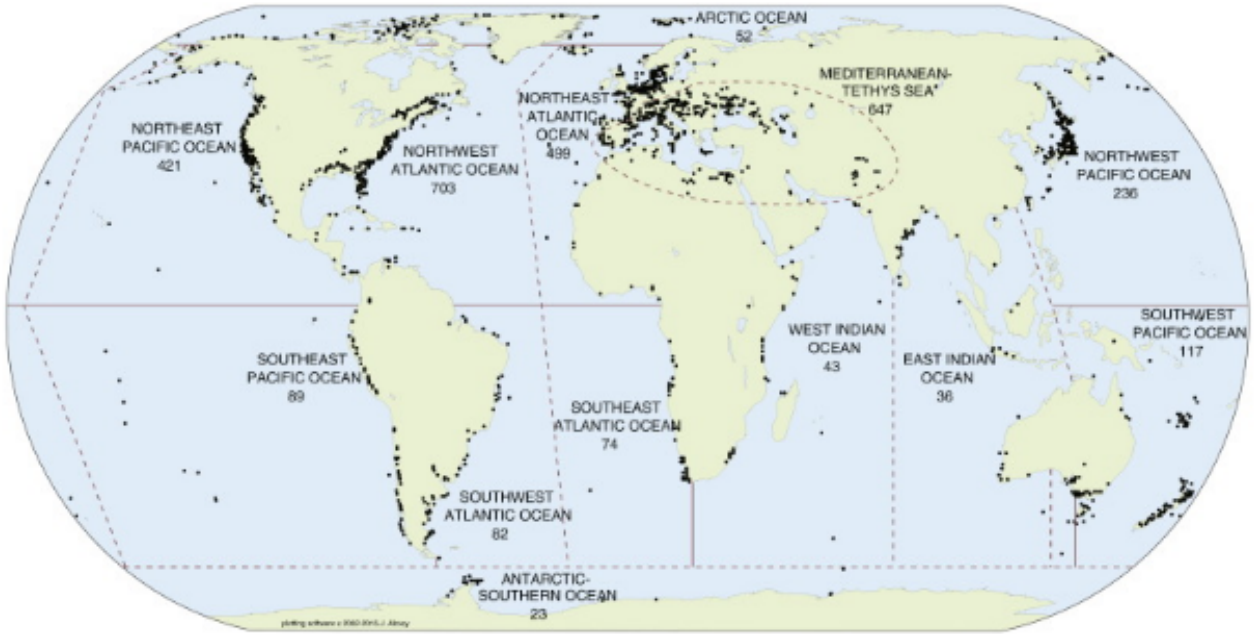


Fig. 1.

Geographical distribution of all 3535 marine mammal collections and division of ocean regions for this study. Each dot represents one collection. Solid lines represent ocean divisions by the International

Imago
Jetstream
+ Digital Library Repository

Imago

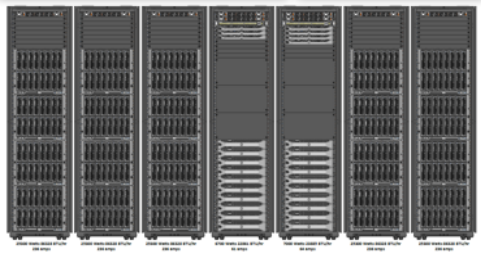
Jetstream

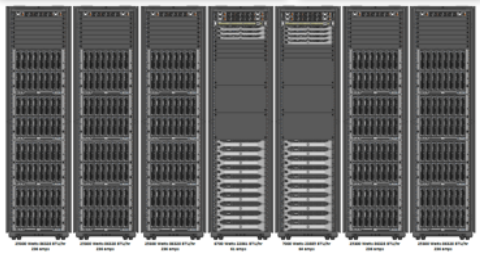
XSEDE

XSEDE Accounting

OpenStack | Ceph

OpenStack | Ceph





Sufia

Hydra | Blacklight | Solr

Fedora Repository 4

You can put YOUR collection online
Via Imago and Jetstream

SUFIA

hydra

DURASPACE™

Broader Impacts

COMBINATION SEARCH

Select State or Province: Utah

Habit (general appearance): Cactus/Succulent

Duration (lifespan): Perennial

Light requirement
- Sun - 6 or more hours of sun per day
- Part shade - 2 to 6 hours of sun per day
- Shade - Less than 2 hours of sun per day

Soil moisture
- Dry - soil does not exhibit visible signs of moisture
- Moist - soil looks and feels damp
- Wet - soil is saturated with water

Bloom Characteristics

Bloom Time: Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec

Bloom Color: White Red Pink Orange Yellow Green Blue Purple Violet Brown Black

Leaf Characteristics

Leaf Arrangement: Alternate Opposite Whorled Fascicled

Leaf Retention: Deciduous Evergreen Semi-evergreen

Size Characteristics

Height: 0-1 ft. 1-3 ft. 3-6 ft. 6-12 ft. 12-36 ft. 36-72 ft. 72-100 ft. More than 100 ft.

Lady Bird Johnson Wildflower Center

Submit combination Search

1 images

1 images

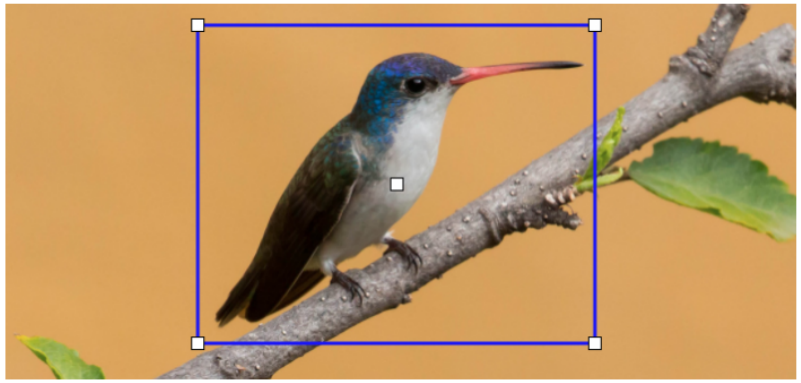
36 images

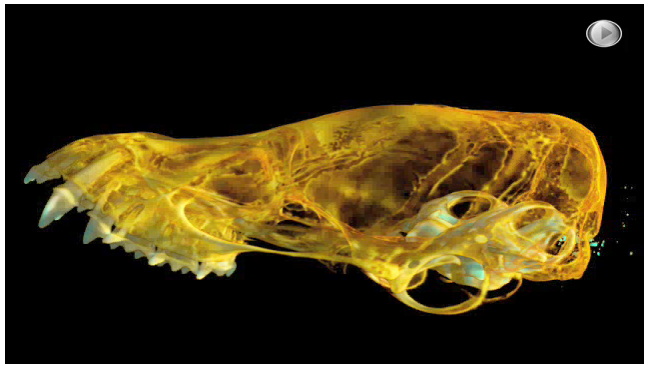
4 images

11 images

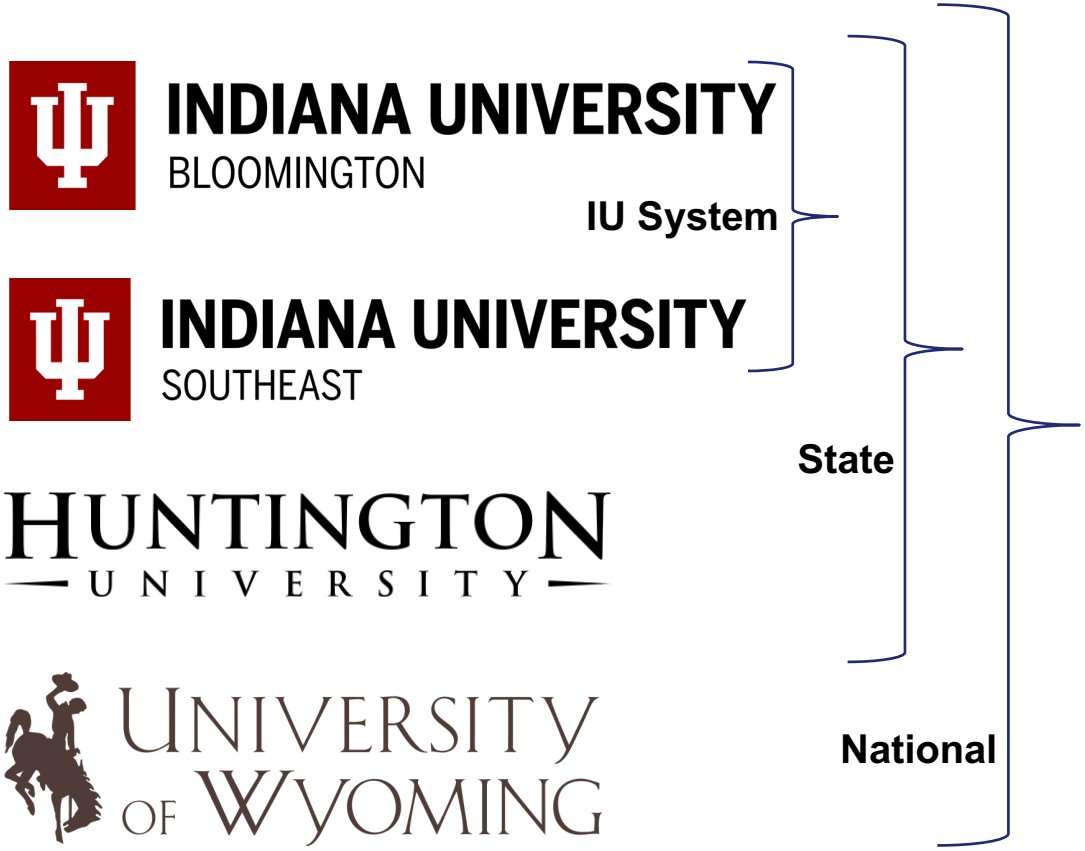
Help us train the next generation of Merlin Bird Photo ID

The Merlin team is hard at work on the next generation of photo ID tools. You can help us train the computer models that will help identify bird photos right on your phone. All you need to do is draw a box around the bird in each photo. Click **Get Started** to give it a try!





Current Partners



Jennifer Laherty, IU Libraries, jlaherty@indiana.edu

Jetstream Timeline…what comes next?

Transitioned to full operations on September 1, 2016

October 2016: **225 XSEDE projects and 650+ users**

Soliciting *Research* allocation requests quarterly plus *Startup* and *Education* allocations *NOW* – including Science Gateways!

Adding services as deemed useful/mature (heat, magnum, trove, manila, etc)

Atmosphere enhancements

Working on partnerships with groups like HubZero

Jetstream
http://jetstream-cloud.org/

NSF funded by the National Science Foundation
Award #ACI-1445604

Resource requests

You can request startup allocations anytime.

You can request allocations for educational use anytime.

Next submission period for large allocations is 15 December 2016 – 15 January 2017.

We are happy to help you prepare a request and create a successful proposal.

You do **not** have to have prior use of Jetstream to be successful.

Where can I get help or learn more?

Production:

Wiki: http://wiki.jetstream-cloud.org

User guides: https://portal.xsede.org/user-guides

XSEDE KB: https://portal.xsede.org/knowledge-base

Email: help@xsede.org

Campus Champions: https://www.xsede.org/campus-champions

Training Videos / Virtual Workshops (TBD)

Jetstream
http://jetstream-cloud.org/

funded by the National Science Foundation
Award #ACI-1445604

Jetstream Partners

funded by the National Science Foundation
Award #ACI-1445604

Questions?

Project website: http://jetstream-cloud.org/

Project email: help@jetstream-cloud.org or help@xsede.org

License Terms

Jetstream is supported by NSF award 1445604 (Craig Stewart, IU, PI)

XSEDE is supported by NSF award 1053575 (John Towns, UIUC, PI)

This research was supported in part by the Indiana University Pervasive Technology Institute, which was established with the assistance of a major award from the Lilly Endowment, Inc. Opinions presented here are those of the author(s) and do not necessarily represent the views of the NSF, IUPTI, IU, or the Lilly Endowment, Inc.

funded by the National Science Foundation
Award #ACI-1445604


Jetstream